"I Always Wanted to See the Night Sky": Blind User Preferences for Sensory Substitution Devices

Giles Hamilton-Fletcher[1], Marianna Obrist[2], Phil Watten[2], Michele Mengucci[3], Jamie Ward[1]

[1]Department of Psychology
University of Sussex, UK
gh71@sussex.ac.uk,
jamiew@sussex.ac.uk

[2]School of Engineering and Informatics
University of Sussex, UK
m.obrist@sussex.ac.uk,
p.l.watten@sussex.ac.uk

[3]LabIO
Lisbon, Portugal
mengucci@gmail.com

ABSTRACT

Sensory Substitution Devices (SSDs) convert visual information into another sensory channel (e.g. sound) to improve the everyday functioning of blind and visually impaired persons (BVIP). However, the range of possible functions and options for translating vision into sound is largely open-ended. To provide constraints on the design of this technology, we interviewed ten BVIPs who were briefly trained in the use of three novel devices that, collectively, showcase a large range of design permutations. The SSDs include the 'Depth-vOICe,' 'Synaestheatre' and 'Creole' that offer high spatial, temporal, and colour resolutions respectively via a variety of sound outputs (electronic tones, instruments, vocals). The participants identified a range of practical concerns in relation to the devices (e.g. curb detection, recognition, mental effort) but also highlighted experiential aspects. This included both curiosity about the visual world (e.g. understanding shades of colour, the shape of cars, seeing the night sky) and the desire for the substituting sound to be responsive to movement of the device and aesthetically engaging.

Author Keywords

Sensory substitution; vision; depth; colour; hearing; sound; blind; visually impaired; experience; design; SSD.

ACM Classification Keywords

Design, experimentation, human factors.

INTRODUCTION

"*Since I lost my vision, I always wanted to see the night sky again, just looking at it... if we could represent the night sky then it could be more interesting for visually impaired and blind people...*"

Visual information, from experiencing breathtaking views to recognising a friend from afar, represents a wealth of life-enhancing information inaccessible to many blind and visually impaired persons (BVIPs). One way this can be provided is through the remaining senses. Devices that encode information normally associated with one sense (e.g. seeing) through another (e.g. hearing) are called Sensory Substitution Devices (SSDs). The overall aim of these devices is to restore some of the functionality afforded by vision for BVIPs. This seemingly clear and simple aim disguises a range of methodological and conceptual problems around the design of such technology. What aspects of vision-related functionality are blind people interested in? What is the optimal way of conveying that information via another sense? What kinds of practical constraints limit the uptake of SSDs in everyday life?

SSDs can be distinguished from other kinds of sensory tools in a number of ways [44]. Unlike the white cane, the information is rich: it represents multiple aspects of the visual environment (e.g. 2D space and luminance). Moreover, unlike recognition aids such as TapTapSee [40], the information operates on visuospatial sensory features (e.g. red, wide, directly in front) rather than semantic ones (e.g. "car"). So while an object recognition system may be able to detect a tree, an SSD is able – in principle – to convey how the tree moves in the breeze, whether its leaves are brown or red, and its distance from the user. However SSDs have to sacrifice some visual features to convey others [18]. For instance, an SSD may sacrifice temporal resolution in order to provide a high spatial resolution [30].

There are a wide variety of approaches to converting visual information into sound and/or touch in the HCI literature [18, 22, 39, 42]. In this paper, we focus on sound and build on the distinction of two approaches: the first is mimicking the experience of external objects emitting sounds (e.g. knowing a bird's location through its chirping) which can be done through replicating our natural hearing mechanisms; the second is to abstractly encode visual dimensions, typically based on intuitive associations between vision and other modalities [36]. These approaches can be combined to varying degrees, use a variety of auditory features (e.g. pitch, loudness, timbre), and be provided to the user either all-at-once or piecemeal. While a variety of approaches have been taken previously [18], there has been little effort to adopt a user-centered approach to these issues. Only aesthetic ratings of two devices have been done previously by BVIPs [3]. Here we seek to give BVIPs various SSDs and an open platform to stimulate discussion on how SSDs can be enhanced to meet their desires and expectations.

CHI'16, May 07-12, 2016, San Jose, CA, USA

DOI:http://dx.doi.org/10.1145/2858036.2858241

In this paper we presented BVIPs with three novel SSDs, representing a large range of possibilities; detailed spatial information (Depth-vOICe); fast spatial information (Synaestheatre); and detailed colour information (Creole). All featuring different sound outputs. Through in-depth interviews with ten BVIPs we gained a richer understanding of the advantages and limitations of these devices for improving everyday life and enriching their experiences. Our contribution to the field is three-fold. First, we advance knowledge on how BVIPs desire to visually experience the world. Second, we discuss the pros and cons of current SSDs for BVIPs. Finally, we demonstrate how future SSDs could be designed to enable BVIPs to reach their desired experiences in daily life and foster long-term adoption.

TURNING VISION INTO SOUND

First we explore how SSDs can turn vision into sound. The most influential visual-to-audio SSD has been the vOICe [www.seeingwithsound.com]. This translates a single 2D greyscale image into sound every second. The visual height gives the pitch of each generated tone, while the brightness drives its loudness. This information is delivered one column at a time, panning from left to right [30]. As such, the following image < would be heard as two electronic tones, one ascending and another descending in pitch, progressing from the left to the right ear. This approach retains a high amount of visual information (see Figure 1).

Figure 1. Progression of visual image into sound by vOICe SSD. Left image shows webcam image, middle image retains only luminance information prior to sound conversion. Right image shows a spectrogram of the resulting sound over one second.

The vOICe has been widely used in scientific research for examining object localisation, recognition, 'visual' acuity, informational content and how this is processed by the brain [4, 5, 7, 8, 17, 31, 33, 37, 38]. Outside of the lab, the vOICe has been downloaded by over 270,000 unique users on android devices alone [1]. Of these users, 21% are reported to be BVIPs [2]. However the vOICe has not been widely adopted by the blind community [5]. Despite its free availability since 1992, only a handful of BVIP users have adopted it for long term use. Considering the potential for visual rehabilitation, why might it be the case that so few appear to persist with the device? While the majority of research has focused on what is possible with the vOICe, little attention has focused on its user experience.

LIVING WITH SENSORY SUBSTITUTION

Reports from long term users of the vOICe device are rare and mostly documented in non-academic sources, despite their potential for revealing key information in user design. Here we summarise experiences reported by three long term users: PF, a late-blind user with minimal light perception and 17 years of vOICe experience; CC, a low vision user with 14 years vOICe experience; and PL, an early-blind user with minimal light perception and 14 years vOICe experience. Below we explore the experiences of vOICe experts, expanding beyond the only journal article on the subject to date [43], to their documentation in blog posts [27], audio talks [14] and popular science articles [10, 41]. From this, we can examine the stumbling blocks and valuable life experiences that these devices enable.

First of all, why did PF pick up the vOICe to begin with? PF comments that "*I wanted a means by which I could identify the colours of clothing in my wardrobe. I did not realise at the time the full potential of the program and how it would eventually so dramatically impact my life*" [14]. For PF, learning how the shape of objects on a scanner bed related to their sounds was a frustrating experience "*...the process was very slow and time consuming. I often stood on the threshold of giving up...*" while the first impactful experience was unexpected "*I shall never forget the shock and joy of glimpsing down my hallway and seeing blinds hanging from the window. Such a simple sight, yet after twenty years of blindness I was overwhelmed... I can see!*" After this PF reports a new found motivation to explore the sounds of the world around her, but even this comes with problems "*It is a struggle to balance... society's acceptance and the wearing of the equipment.*" PF later took part in an experiment that temporarily disrupted her ability to use the vOICe [31], it ended up being a hugely emotive experience. "*[The] world started getting darker, like someone slowly turning down the lights... I couldn't see anything... I wanted to cry because I thought they broke my sight*" [41]. For PF, the offer of practical colour discrimination lead into sensory substitution, while the later valuable experiential gains were unanticipated. With time PF reported visual colour, motion, and depth perception from the vOICe [43].

CC also reports a long effortful learning period "*It took three months for me to learn enough so that I didn't have to consciously concentrate on it.*" However over time expertise with the vOICe has created a smoother experience that co-exists with her remaining visual perception "*...you see the blurry sort of dark version with my eyes and a very crisp clear... version with the camera of the phone.*" Overall CC uses the vOICe for locating objects like coffee mugs and adding to her travelling "*I've sailed across the English Channel and across the North Sea... the lights on the land were faint but the vOICe could pick them up*" [41].

For PL the effort required to use the vOICe has decreased over time but isn't gone. "*I do not think consciously about the rules anymore, I sense a scene and then break it down into shapes... if I encounter something really new, then I know the three basic rules and try to make sense of it*" [27]. Now he uses his expertise with the vOICe to take photos, notably based on the auditory signal rather than the visual. "*Scenes with varied elements sound interesting, for example, a bridge with lots of cables or a bright natural* *setting with streams and rocks... It's like an image orchestra*" [10]. For PL, the use of the vOICe appears an important aspect of his experience "*I experience so much more... I was looking at all the vehicles and at the walls on the side of the road... words do not convey visual information, you need to experience it.*"

PF, CC and PL are users that have pushed past the demands required to effectively use the vOICe, but represent only a fraction of total BVIPs who have experienced it. There appear to be substantial experiential gains from long term use. So how can designers create SSDs that also appeal to initial users, and provide richer experiences earlier on? First of all we need to re-examine options beyond the vOICe.

DESIGNING SENSORY SUBSTITUTION DEVICES

Here we highlight a variety of visual and audio possibilities for SSD design, before considering the design of our SSDs.

Visual possibilities

By default, the vOICe translates a 2D greyscale image every second. Is luminance perception the most desired visual characteristic that SSDs can provide to BVIPs? Some users may already have light perception that borders on the same resolution as SSDs [17, 38]. Instead, users may be interested in additional information for low-luminance environments. In experiments using the vOICe, silhouettes of objects are chosen to simplify information [5, 8, 11, 23, 33, 35]. Similarly, novice users perform best when objects rather than luminance produce the loudest sounds [8].

One alternative is using 3D spatial information (see Figure 2). Depth maps are able to simplify complex scenes by prioritising and segmenting nearby objects from their more distant backgrounds. As a result key objects become simple shapes, similar to those used during training with the vOICe, yet immediately practical to novice users. This is particularly relevant now, with the advent of depth sensing technology on mobile phones [20, 32].

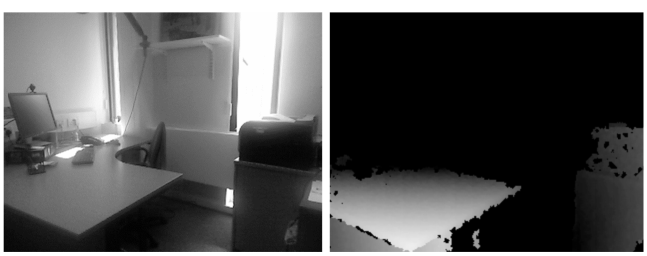

Figure 2. (Left image) Luminance provides a lot of information, but much of this is unnecessary to the end user. (Right image) However, depth images provide a simplified and prioritised image.

PF's initial fascination with SSD technology came from its potential to sense the colour of her clothing. As such, information about hue and saturation may be desirable to BVIPs. How complex do BVIPs desire colour information to be? Is knowing that an object is red enough, or are users interested in more subtle shades of colour? How will BVIPs desire using this beyond discriminating clothing?

Audio possibilities

The design choices of the vOICe fit well with intuitive tendencies about how different sensory dimensions are related – so-called cross-modal correspondences [36]. For instance, high pitch is intuitively linked to high spatial positions, and loudness is intuitively linked to brightness [12, 28]. However, there are many other possibilities for representing visual space in the auditory domain. How do users respond to combinations of pitch and loudness to receive their visual information? How does the introduction of richer timbral sounds such as vocals or instruments influence opinions? Do users find particular sounds easier to understand objects or the visual scene as a whole?

PF, CC and PL report that using the vOICe initially required a lot of effort. This effort can be reduced by using natural hearing mechanisms that avoid conscious deduction. Natural hearing is able to discriminate the location of sound sources via differences in interaural timing, interaural intensity, and distortions in the sound signal created by the head and ears. These in combination are described using a head-related transfer function (HRTF). Related, BVIPs tend to have enhanced sound localisation in the horizontal, but not vertical, plane [24, 34]. So HRTF could provide an easy-to-learn auditory coding of visuospatial information.

Finally there is the question of if BVIPs prefer hearing all available information at once, or in a piecemeal fashion like the vOICe? Do the potential advantages of piecemeal presentation (e.g. resolution) outweigh the fact that the signal is slow to update when the device is moved?

Other features

Beyond sensory information, is there any other information BVIPs want SSDs to provide? What other characteristics could make SSDs more desirable? How do users want to control SSDs? By understanding these questions from the end user perspective, it may be easier to foster long term adoption by meeting the needs of the BVI community. In our research we focused on the exploration of blind users preferences for SSDs by presenting them with a wide range of device options and experiences, more specifically three

Device	Information	Output	Sounds	Spatial resolution	Information at a time	Temporal resolution
Depth-vOICe	3D space	Bone conductance	Tones	176*64	1 column	1.5 seconds
Synaestheatre	3D space	Headphones	Instruments	13*7	All locations	Instant
Creole	2D colour	Speakers	Tones / Vocals	1080*720	Single pixel	Instant

Table 1. Comparison of three sensory substitution devices (SSDs) representing a large parameter space of possibilities for enabling blind or visually impaired people to see the world through sound. All three SSDs were used in the interviews with BVIPs.

new and easy-to-use SSDs (Table 1). These devices are introduced in the following section before the study set up.

THREE NOVEL SENSORY SUBSTITUTION DEVICES

Three new SSDs were presented to participants that collectively showcase a large range of visual and auditory configurations (see Table 1). The Depth-vOICe scans through a single image over 1.5 seconds to provide highly detailed spatial information, enabling shape recognition. This is delivered as a series of electronic tones, where objects present in a depth image (see Figure 2) are scanned across one column at a time, with the pitch of these tones denoting the object's vertical position, left-right panning giving its horizontal position and finally loudness gives the closeness of the object. By contrast, the Synaestheatre gives a lower spatial resolution (13*7) in order to instantly provide information about all the locations at once, this fast temporal resolution makes it suitable for detecting motion. This is presented through instrumental sounds that vary in their pitch, panning and loudness according to the vertical, horizontal and distance of objects captured in a depth image. Finally, the Creole sacrifices wide scale spatial information for highly detailed, localised colour information. Individual pixels are selected using a tablet and stylus and their colour information is delivered through a combination of tones and vocal sounds.

These three devices are best suited to interrogating different elements of an image, e.g., when viewing a car, the Depth-vOICe's high spatial resolution makes it suitable for identifying the car's shape; the Synaestheatre's fast refresh rate is good for identifying its speed; while the Creole can identify the car's colour. Overall users have access to a wide range visual sensory characteristics and auditory feedback methods, designed to facilitate dialogue about what desirable SSDs might look like in the future and how this may feature in BVIP's lives.

SSD#1 – The Depth-vOICe

To provide practical, easy to use SSDs, we used the vOICe to sonify a luminance-depth map as this process simplifies a complex visual environment into basic silhouetted shapes. A 80*64 pixel depth map is provided by a custom C# application which receives depth stream information from a Xbox 360 Microsoft Kinect camera. This depth information is visually presented using luminance ramped in a linear fashion between the set minimum (white) and maximum (black) distances. Minimum distance was set at 0.8 metres and maximum distance could be varied between 1 and 3 metres in half metre increments. The depth image was turned into sound using the vOICe's sonify active GUI client command. Following initial consultation with a blind user, the following non-default options were chosen to increase ease of use. The vOICe now takes an image every 1.5 seconds to be scanned through and the scan progressed back and forth between the left and right sides with 'clicks' occurring on the left and right end points. Options for pitch range (500-5000Hz), panning and loudness remained default. The sounds were delivered using bone conductance headphones (AfterShokz). This stimulates the inner ear via bone vibrations and avoids masking sounds received by the ear [15]. With minimal training, one blind user was able to locate objects in an open environment and identify the shape of a person. This is the first piece of research to showcase the Depth-vOICe.

SSD #2 – Synaestheatre

One alternative to the vOICe's approach is to replicate the experience of the external objects themselves emitting sounds. For example, if a bird chirps while flying around you, the sounds heard by each ear will vary in a predictable way. This effect can be modeled and recreated using HRTF, which creates the sensation of individual sounds originating from a specific location. HRTF has been used before in some devices to represent horizontal information [6, 16], but have not been evaluated from a user perspective for SSDs relative to alternative methods. Replicating natural hearing should provide immediate, low effort spatialisation.

The Synaestheatre SSD translates a Microsoft Kinect Camera depth stream into 91 depth points (13*7 pixels), each of these points control the volume of one of 91 pre-recorded sounds to be played when an object enters between its minimum and maximum depth values. This is run in Pure Data using Freenect drivers. The pre-recorded sounds consist of an A3 pitch (220Hz) banjo sound lasting 1 second, recorded by the Royal Philharmonic Orchestra [www.philharmonia.co.uk/]. This sound had its pitch altered through the Audacity's change pitch function [http://audacityteam.org/] to produce 7 mono sounds at 110, 165, 220, 330, 440, 660 and 880Hz. These make up the 7 vertical positions. To fill out the horizontal positions, each of these were turned into stereo sounds with varying spatial positions produced through the Panorama plug-in by WaveArts [http://wavearts.com/] for the Reaper audio studio [http://www.reaper.fm/]. The panorama plug-in created interaural timing and intensity differences for each of the sounds. The sound source was 2 metres from the listener's head position, at 13 horizontal locations ranging between -90 to +90°s in 15° increments using the MIT-Kemar HRTF dummy head model [http://sound.media.mit.edu/resources/KEMAR.html]. In combination, the pitch and HRTF variations produce 91 pre-recorded sounds, each of which lasts 1 second.

Spatial information is given to the user from all 91 depth points simultaneously, with the sounds synced together with a one second tempo. The temporal resolution is faster than the sound sample because the loudness of each of these sounds is instantly varied by changes in spatial information even if the sound sample has already started playing. Distances under 0.8 metres are played at maximum loudness, and linearly reduce in volume to silence when the set maximum distance is exceeded in the pixel's location. The maximum distance was set to 2 metres due to the interview room size. As such, a single object would sound like a banjo constantly playing that can shift instantly in space, pitch or loudness by varying the objects' location. The Synaestheatre has not been showcased before.

SSD#3 – The Creole

The Creole represents 2D colour information that can be explored serially using a stylus on a tablet (see Figure 3).

Figure 3. The Creole SSD. Left to right: A 2D colour image is loaded and explored by the user using a stylus on a tablet device. The stylus position on the tablet corresponds to the position in the image (see green dot). The colour at that spatial position is turned into sound. Over time the user can understand the whole image. The rightmost image is a picture drawn by a user of the device in response to hearing the image on the left.

This device is based on a previous SSD used to explore the use of correspondences for colour processing [19, 20]. A 2D colour image is loaded into a Jscript program running in Google Chrome which selects a single pixel underneath the mouse cursor and systematically translates this into sound. The pixel's RGB values are transformed into CIE LUV colour space. From this, proportions of white, grey, black, red, green, yellow and blue are taken. Each of these colours is associated with a specific sound as detailed in previous publications [19, 20]. CIE LUV colour space is perceptually uniform, so the perceptual distance between colours relates to the perceptual distance between sounds heard. Participants used a Bamboo tablet and stylus to interact with the SSD. The Creole was featured in one previous study using sighted participants [19, 20].

STUDY SET UP

Ten participants with blindness or visual-impairment were recruited from the University of Sussex, UK and the Blatchington Court Trust. We included respondents who "cannot rely on vision in day to day life." More details on each participant's visual impairment level and background can be found in Table 2. This was approved by the University Cross-Schools Research Ethics Committee.

Participants were reimbursed $15 per hour interviewed (typically 2-4 hours) alongside travel expenses. The interviews were audio recorded for analysis purposes. The key interview questions were as follows: (1) "Have you had any experience of vision? Has this changed? What do you experience now?" (2) "What are the biggest difficulties to overcome in [navigation / locating objects / discriminating objects]." Interviewees are then given a hands-on session with the SSDs and subsequently asked (3) "what are your initial impressions of the devices" and (4) "what are the most important improvements that could be made?"

Midway through the interview, participants were able to try out each of the three SSDs in turn, typically taking 20-30 minutes with each device. All devices were used within the same room as the interviews. The first two 3D sensing SSDs (the Depth-vOICe and Synaestheatre) had their visual field-of-view (FOV) and maximum distances explained to users. Explanations were assisted by using tactile aids consisting of card rectangles with raised cut-outs representing the device's FOV and objects respectively. In conjunction with this, the device's translation of horizontal, vertical and depth positions were explained to users (e.g. a single object in the top-left creates a high-pitched sound on the left), typically taking 5 minutes. Users tried the devices themselves in simple scenarios, consisting of the presence of one object (head-height cardboard circle 30cm in diameter supported by a thin stick), which could be listened to in any location in the device's FOV. Then users listened to the experimenter's profile, either neutral or with arms outstretched in different directions, as well as the experimenter moving across the device's FOV, both left-to-right and away-towards. Finally users had free exploration of the environment for 10-20 minutes, listening to chairs, tables, shelves, walls, windows and the ceiling. Finally, since the Creole uses pre-prepared images, a different approach to training was taken. The principles of the device were verbally explained and users were then asked to find three hidden coloured circles on the tablet moving the stylus and listening for sound changes (indicating colour changes). After this, users were able to explore a 2nd image consisting of a palette of colours. Users were encouraged to gradually go through the palette, while the sound changes were explained to users. Overall we followed an open and

Participant	Age	Sex	Background on Visual Impairment	Current Vision (Self Reported)
LP*	23	F	23 years no vision (retinal cancer)	None (congenitally blind)
VG*	20	F	20 years no vision (Leber's congenital amaurosis)	None (congenitally blind)
DHO	25	M	3 years low vision (retinoblastoma), 22 years no vision (retinal detachment)	None
DHA	21	M	14 years low vision, 7 years no vision (detached retinas)	None
GS	44	F	17 years low vision, 27 years residual light perception (unknown cause)	Residual light perception
HS	24	F	24 years deterioration to residual light perception (genetic disorder)	Residual light perception
DB	23	M	23 years low vision, 8 years no vision right eye (retinal detachment)	1-2% visual acuity in left eye
MC	22	M	22 years low vision (Leber's congenital amaurosis)	5-7% visual acuity in both eyes
DW	32	M	32 years low vision (Leber's congenital amaurosis)	15-20% visual acuity in both eyes
CJ	70	M	62 years normal, 8 years low vision (age-related macular degeneration).	Peripheral vision only

Table 2. Overview of participants' demographics and different levels of visual impairment. M = male, F = female, * = No visual experience.

exploratory approach to guide the exploration of the devices. Users were encouraged to ask questions to aid learning. Please see the supplementary material for a video showcasing the interaction for all three SSDs.

After each device, participants were asked about their impressions and concerns for these types of SSDs, as well as what they would want to use it for, if at all. Finally they were asked whether SSDs were of interest to them and what features would be ideal for their needs. Participants were also asked if they think they would use an SSD in future, on a 0-10 scale (where 10 is "definitely").

RESULTS

The four main topics covered in the interview guided the analysis of the data. Key themes were summarized based on repeated readings of participant statements and discussions between the main author (interviewer) and the co-authors of this paper. Relevant information extracted from the interviews was further discussed with respect to the overall goal of the research to identify experience-relevant design opportunities for sensory substation devices.

The main results are presented along four main themes: (1) Visual aspiration, (2) Practical problems, (3) Experiences with SSDs (i.e. Depth-vOICe, Synaestheatre and Creole) and (4) Desires for future SSDs. Each theme is introduced in the following section and exemplified through participants' quotes.

Theme 1: Visual aspirations

DHA who is currently studying physics expressed an interest in visual information relevant to his work "*Since I lost my vision, I always wanted to see the night sky again, just looking at it... if we could represent the night sky then it could be more interesting for visually impaired and blind people... it would really open-up astronomy as a scientific field*." Some desired visual experiences go beyond immediate practical concerns and extend to new experiences as well as helping interactions with others "*They talk about like how beautiful Venus and Jupiter look like... which I don't remember looking at... I would absolutely love to see that... it would definitely open up conversations*." Opening up possibilities for interaction was a re-occurring topic in the interviews. MC "*When friends meet they might say 'oh that's a nice new jacket' ...that's the sort of thing I could miss. Just in my personality, that's something that I would recognise and comment on, if only I could see it.*" CJ, for example emphasized that "*...the most important thing you'll ever see is a person smile... [it would sound] softly if it's a nice sweet smile, raucously if it's an outrageous laugh or smile...*" The subtly of such emotional expressions are not covered in current SSDs, but is a desire, especially for people in the transition of losing their vision.

Theme 2: Practical problems

Beyond aspirations of attaining visual information, practical everyday challenges were widely mentioned with respect to navigation, localisation and discrimination.

A huge variety of navigational issues were identified. DHA "*You need somebody to teach you the routes, so if you charted somewhere you've never been before, well unless you're really really brave, there is little chance you can find a place you want to find. It's a huge drawback in navigation that you need somebody to teach you the route, which occasionally costs quite a bit of money, very time consuming and you need to memorise everything.*" DHO expands on this "*Sometimes the person who's helping you might not really have much to say... and you feel a bit awkward.*" MC who has low vision mentions "*It takes me a great deal of concentration to navigate at night.*" DW "*The biggest problem when navigating anywhere is other people*." Whereas VG* was most concerned about ground level changes "*Having lots of steps as well... quite tricky, especially when I don't know how many.*"

Most participants use routine to locate and place back objects in daily life. A major difficulty results from others using these objects and not following the same routine. DB "*A lot of it is guessing where people have put it, if you live with someone and they've moved something, at least you've lived with them and you know their habits ...if it's a stranger it could be anywhere.*" Objects that do not naturally enter daily experience can limit the options available. DHA "*Not knowing that there is a bin in a certain area... 10 metres away there might be a bin I never knew about... you really only find things you are looking for.*"

Most participants use touch to discriminate between objects, alternatively software (TapTapSee), or video calling (Facetime or 'Be My Eyes') were used, but had downsides. LP* "*We have a big box that's got all the… little herb jars, they're all the same! …things like TapTapSee would come in, but, it's time! It takes a while to take a photo of each individual herb jar, if you've got ten of them and you only want one. That might take a while. So it's frustration in terms of how long that might take.*" Not knowing what an object is, or the state that it is in can be hazardous. DHA "*Sometimes you cannot be sure if you can touch it or not.... you want to check if it's a hob... you cannot be 100% sure that it's not turned on... so you have to be careful what you touch*."

There are a wide variety of problems that BVIPs experience, so next we explore if current SSDs offer information that users could find useful or interesting.

Theme 3: Experiences with the SSDs

Based on participant's free exploration with each of the three novel SSDs, we summarise key impressions below.

#1 Depth-vOICe

Many of the positive comments relating to this device came from the fact that they (unexpectedly) could appreciate the perspective of objects and the relative positioning of objects. DB [listening to objects on a shelf] "*They almost have their own unique signature... that would kind of be helpful though, to know whether you are straight with* *something or not...* [listens to a coat, then cane, hanging on a door] *yeah, it's there, definitely, that's very interesting… to think that it can pick something that small up as well.*" DW "*it's telling me that the shelves are here, that the left hand side of the shelves is much further away than the right... complicated but easy to learn...*"

Many participants expressed optimism on effectively using the device or combining its information with their residual vision. MC "*It's impressive you can use the loudness and pitch to make a construct that actually isn't difficult to calculate what it means… if you're walking down the street… you're gonna hear if the car's parked in your way, and just by the breadth of sound from left to right you'll be able to have a good guess at what sort of shape it is.*"

Negative comments were made about the auditory aesthetics of the device, and the cognitive effort required. DHA "*I don't expect the room to play Mozart for me... but I don't expect it to make some crazy electronic frequencies that give you a headache after 5 minutes... just something pleasant I can listen to for 30 minutes... something more instrumental, more natural.*" GS predicts multitasking problems "*...that's probably why I gave up on the 'seeing with sound' thing because it takes too much out of you, especially when you are moving, you can't possibly use a cane and at the same time concentrate on these noises and thereby trying to figure out what object is in front of you.*"

Participants saw the device enhancing navigation, locating new objects and listening to interesting shapes. DHO "*Outside ...that's when you could really gain useful things because your scenery will change a lot... you could get a little bit of pleasure even if you knew it was a route you did often, I walk across this field... getting to see it for the first time or getting to actually go beyond what you just pick up from the cane and the path that you stick to...*" HS "*I'd probably use it for home use... if I'm navigating around my flat, seeing how close I am to things... I would use it to maybe navigate outside on my own, which I don't do at the moment.*" DW "*I would use it at work for looking for objects on the stage or on the floor...*" DHA "*Having a look at some architecture, like the Eiffel Tower.*"

Overall participants expressed interest in the location, perspective and shape of objects while navigating, however the poor aesthetics, difficulty of object identification and effort required, limits its application for many users.

#2 Synaesthеatre

The most positive impressions of this device related to the fast feedback, low-effort localisation and auditory aesthetics. DHA "*One of the coolest things is that it gives instant feedback, so as I turned the camera the sound just instantly changed... it was just all so much easier to distinguish between left and right because as I turned the sound just moved to my left ear or right ear... this one is significantly easier to interpret... As I was hearing you moving from right to left... it was clear when I'm hearing a stationary thing or a moving thing.*" Using various pitches for an authentic musical instrument made interactions more fun for some users. CJ "*It's a lovely sound... I could probably make music... this would be great fun, you could enjoy this, you could design rooms which would play you [music]… it feels like fun.*" MC "*The tones it makes are almost enchanting…*" The all-at-once approach seemed to be easier to understand in some ways. DB "*It's almost what I wanted the first device to be… but I didn't know it could work… you're getting a flash of the entire visual field… I find it so much easier to process the same information.*" DHA "*This is the first time I've used it and I think I've already learned more... than the previous one.*" Furthermore, the audio made even simple interactions fun. DB "[puts hand in front of the camera] *oh dude, I can hear, look ok, index finger, thumb* [listens to each in turn]... *that is fantastic… that is really perceptive, that is amazing*."

The negative impressions mostly related to identifying shape information and object identification. GS "*When it comes to sweeping from one side to the other, the first device is better... to indicate the width of an object...*" DHO "*I could see it being potentially more of a problem if you are trying to identify an object... you can't immediately find out what something is.*" Some users wanted the information and tempo even faster. CJ "*I could cope with faster... it needs to be the speed at which my eyes... can handle the information, and that's virtually instantaneously.*" In addition, mismatches for the representation of an object in real space created instant confusion. DB "*Things start suddenly sounding behind you... if there are two objects either side... where it's not representative of where things are in natural space it throws you*" this likely reflects the 'cone of confusion' problem resulting from ambiguous localisation cues. LP* mentions that quieter elements of the sound signal may be missed "*if there is something really quiet going at the same time, you're probably not going to be aware that it's there.*" Resolving these issues may involve adding timing differences, using smaller panning ranges and natural distance-hearing models. LP* "*If you could combine the two and be able to choose between having your sweep and having the focus... If I was going to assess a room, I'd do the sweep first and then that one.*"

Participants primarily saw themselves adding to navigation with the Synaestheatre device. DHO "*it would be just giving you that reassurance ... if I remember that that thing is two metres to my left that means I'm staying on the right path...*" The aesthetics allowed relaxing navigation as well. DHO "*If I was feeling... headachy... but I still fancied the walk.*" Some users see it as a replacement for their prior navigation techniques "*You walk along... pedestrianised roads and if people don't take care of their hedges or their trees or whatever, sometimes you're having to constantly shield your face with your hand... and it sucks... that* [Synaestheatre] *would be a very very handy pain free way of avoiding obstacles like that.*" Users lamented the lack of shape information in new environments however. DHO "*I'd* *like to take the* [Depth-vOICe] *on more of a place that I'd never like been to, after a bit of learning, because I could potentially gain more about the type of object... While I would still be quite rubbish between the difference between a square and a circle, I could eventually learn that.*"

Overall the Synaesthe atre device was praised for its low-effort, fast and 'fun' approach to providing spatial information, but the lack of shape recognition abilities impacted its perceived utility in new environments.

#3 Creole

Users primarily responded positively to the aesthetic and educational possibilities of the device. DHO "*...when people would ask me about the [luminance] vOICe first, some people, from a sighted point of view would be like 'yeah you could see what I looked like' and I'd say like well to be realistic I don't think I'd ever gain anything from doing so... but something like this could actually put some pleasure into it, giving it a defined feeling...*" DB "*That's helping me already learn about how colour works... there are massive holes in my knowledge of how colours and tones and things like that work and if it's represented like... ok, bright blue has a lot of white in it, stick some white [sound] in that, simple, it's so easy and so intuitive like that, I mean it's a new concept for me, but it makes sense... it's all well knowing that there is light, but knowing how that interacts and works with other things could be really really useful.*" MC "*I could never follow… colours at school. It was one of the reasons I didn't like primary school… something like this… a visually impaired student could… learn about colour with just sound if they can't see at all… so they don't feel left out.*"

Negative aspects about the Creole concerned the lack of feedback about colour names, potential training and specific sounds. MC "*You could even have the colour on the tablet screen as you explore it, so it's explained to you… it could be read aloud.*" HS "*Making sure that if someone was to use it, they wouldn't need too much training on how to use it... maybe incorporate some speech.*" The specific sounds themselves had a variety of criticisms. CJ "*It would take a bit of care to get the sounds right... might have to increase the scale of differentiation between the different colours.*" Some desired an even higher resolution of colour than provided here. DB "*It didn't feel specific enough for me when we were looking at saturation... I didn't really detect... more of those... minute changes.*"

The ability to customise sound-colour combinations was a common desire. CJ "*Select your own sounds to colours...*" DHO "*I actively dislike some of the noises... I don't know if the sounds need to be changed...*" Some users saw the device as a skill to be mastered. CJ "*To get the best, you'd have to work at it for years, the same way you'd use a piano or musical instrument, it's that sort of experience... very difficult but extraordinarily rewarding.*" Distinctiveness and fast feedback on the sounds were key topics. GS "*I didn't need much training for me to recognise how one colour popped out from the other based on the sound... it changed in real time which is good.*"

Detailed colour identification across a range of clothes was a commonly desired use for the device. GS "*if I were to put it on my trousers and it were to make different noises... if there are several colours in a dress then usually* [verbal colour detectors] *don't get it right and especially if there are bits of green and white and yellow and blue on a dress, it only tells you the predominant colour... most blind people use colour detectors to look at dress colours and the colour of the clothing they are wearing, the colour of the socks...*" Congenitally blind participant LP* said "*I think it's more in terms of practicality, knowing what colour something is can be helpful, but that's the only reason I would ever use it.*"

Its perceived utility also expanded to science and art. DHA "*A pie chart... that could be quite useful in statistics for instance... At the moment I'm dealing with what colours to use for the website, yeah it would be nice to have a palette of colours to choose from and pick a colour based on the sound.*" DHA also noted the emotional content of the colours as desirable "*I want colours that are really calming, peaceful.*" DW "*it could be used in photo editing... or painting...*" DHO "*I'm a big fan of reminiscing and that, and even if they were old family albums... the colours of objects which were bygone and no longer around [would interest me]*." MC "*I do really like... to design things on the computer... presentations, leaflets, posters... I could use that to independently work with colour, rather than needing to grab a sighted person...*"

Overall there was a desire for very high resolution colour information that was fast and aesthetically pleasing, alongside the verbal labels for confirmation.

Theme 4: Desires for future SSDs.

For spatial and colour SSDs a few key factors continually reoccurred throughout the interviews.

The first factor appeared to be the sound quality of the SSD, with preferences for high quality sounds that were fun to engage with. DHO "*I liked the orchestral ones ...the electronic ones are never nice for too long...*" GS "*...as long as it's not too harsh.*" Some rated this as highly as practicality. DHA "*It's probably equal importance that it's also aesthetic and pleasing to use then just functionality.*"

Some users took issue with low aesthetic concerns for assistive devices in general. DHA "*Most of the assistive technologies just focus on functionality... this machine can tell you what colour your shirt is, but it doesn't put any emphasis on aesthetics... for the sighted community we put so much effort on design, it sounds so fancy and looks so good... blind people also require these things... it's not sight specific, it's human specific.*" MC "*I'm proud to show technology that I'm using to people, and feel that it makes me cool and sophisticated and forward looking... using technology that you're comfortable with and empowers you I think is a very positive thing.*"

Without exception, bone conductance was the preferred method of receiving the sounds, specifically to not override natural hearing. MC "*Bone conduction is the perfect way to receive this input of the sound without impeding your other senses at all while you're out.*" GS notes that it's still not a perfect solution "*...[the sounds] can still distract you.*"

SSDs were particularly desirable as a way to control whether the user appeared 'blind' to wider society. LP* noted that "*People... don't want to look overly blind and stick out like a sore thumb, that's going to be important.*" CJ "*It has to be totally cosmetic, I don't want to look different... I would use it instead of a long white stick.*" DB "*...you are now in the age of inclusive technology and it has to look as [inconspicuous] as possible.*"

The mental effort involved in using SSDs was a major concern. This attribute was frequently balanced against the effort required for their current approaches to the situation. DHO "*That was all a big debate with the [luminance] vOICe as well, they're saying, oh it's going to take you know 3 hours a day... you're going to wear yourself out.*" MC "*Mobility wise at night... you can't do... multi-tasking while you're on the move, like talk on the phone or dictate something... because so much more of my focus is taken up by keeping track in my mind of where I am and using the long cane to find that one point of reference after another… while there might be a learning curve, ultimately [SSDs] might be greater, because it's more efficient and thus gives you back the ability to do... multitasking on the move*."

Many users desired visual information about changes in ground level via the SSD. DW "*Can you program it to identify a step going down?*" LP* "*If you're walking down a straight corridor...you're alright, [but its] not going to help you if there is a staircase going down...*"

For nearly all participants, having any future SSD would only be acceptable if it was hands-free. DHA "*Not something I would have to hold... you have to hold your white cane in one hand, shopping bags in another, then your mobile phone in your third hand, so that's a bit problematic*." LP* "*I think it would be good to have the option, if it was a handheld one, then you could get a strap for it, to mount it on something else… body harness and a head harness, to have that choice.*"

Of all the information that could be provided, spatial tended to be most desired. DHO "*I think that's the top of all of the things that the sensory substitution could provide, spatial information is the most relevant...*" as well as be compatible with higher level navigational information. Building up expectations for tactile interaction was also desired. DB "*I'd like textural information as well, it would be nice to know, if that set of double doors is wooden or not...*" High resolution colour information was of interest to many participants with visual experience. MC "*It's got to go beyond the basic colours, there's too much colour out there to just to say 'it's green,' or 'it's blue,' the sky is blue most of the time, but it can be a bright gorgeous blue, or pale, or dark and ominous... it's got to go beyond basics.*"

Object recognition was prioritised above sensory information for some blind users. LP* "*so it's telling me there is an object… I don't know that's a filing cabinet, so I would have to go over to it and feel it myself to know what it is… maybe a way of... being able to identify it.*" GS "*...it's probably easier to kind of program an audio software to make audio announcements, you know, if it sees a shelf as shelf... the fact that there is an object, that's not too difficult to find out... that's really not the problem.*"

There was large agreement on the use of buttons or phone-swipes to assess options above verbal commands due to environmental noise. Most users wanted the technology to feature on their phone. CJ "*I don't want to carry anything else around... I've got enough stuff to carry around, I want it combined*." However for any phone-based SSDs, the phone itself cannot be exposed to danger. DW "*If you're having a look around with your phone... that's going to be so easy for someone to just take it out of your hand... I think blind people would be robbed considerably... as a cabled [camera] I think that could work... keeping the phone safe I think has got to be a priority.*"

In summary, the interviews revealed a core set of future considerations relating to: sound quality and aesthetics; social considerations; and practical ones (e.g. not hand-held). The final question asked about future likely use of an SSD (0-10 scale) and the overall response was positive with a mean agreement of 8.4 (SD = 1.26), despite the reservations related to the current devices.

DISCUSSION AND DESIGN INSPIRATION

The interview findings helped identify common concerns with SSDs and revealed a strong desire for appealing aesthetics (e.g. pleasant sounds) and to facilitate positive social interactions (blending in, having something 'cool'), as well as having practical benefit. In the next section we first contrast the main feedback on the three devices and then discuss more specific implications for SSD design.

Beyond existing sensory substitution devices

Considering the three devices we examined, the Depth-vOICe was praised for segmenting and locating objects, as well as for perspective (e.g. slanting shelves). Users with prior experience of the vOICe, based on luminance-to-loudness, found depth-related signals to be more intuitive. While the Depth-vOICe has a slow sensory-motor loop (1.5 seconds) this was not the case with Synaesthetre. For this device, participants liked the fact that it responded quickly to visual changes. The auditory characteristics of this device, based on a musical instrument, were the most appealing of the three devices. The Synaestheatre's use of naturalistic sounds and localisation (based on HRTF) is an important design feature moving forwards. The Creole's conveyance of detailed colour information was strongly desired, and could be incorporated into other SSDs [3]. BVIPs expressed an interest both in knowing the colour ("red") and the shades and inter-relations of colours (e.g. understanding that light blue is similar to white).

Using these three devices has further revealed the desire for task-specific adjustments through a fine degree of control for the end user. Adjustability to individual preferences and circumstances need to gain more attention by SSD designers to make devices more desirable. Since ultimately tradeoffs need to be made (e.g. spatial detail vs colour detail), these could be dynamic, so that the user selects their own priorities for the task at hand. This way an SSD may become flexible across multiple scenarios (e.g. admiring the night sky, or finding that red shirt).

Technology that empowers problem solving

Users expressed an interest in technology facilitating practical solutions. For navigation, users identified a variety of obstacles (people, unexpected objects and curbs) that are not currently prioritised by existing SSDs in an intuitive way. This focus is backed up by a recent survey by the Royal National Institute for the Blind, which found that 95% of respondents had experienced a collision outside of the home in the last 3 months leading to physical injury and a loss of confidence [45]. Reliable spatial information was highly desired for users currently unable to effectively deal with low lighting conditions. Sensing changes in space may enable the identification of landmarks to guide navigation. This in turn provides confidence and expands the option of independent navigation, giving BVIPs further control and empowerment over their lives and experiences.

Interviewees expressed interest in several areas currently being investigated in the wider HCI literature. In regards to guided navigation [45] with the Synaestheatre DB says "*If I was following someone… I've got a real-time representation of whether they are going further away from me, or closer or whether I have to walk faster…*" Likewise, DB mentions "*Trying to find my way across very large open spaces without veering too much can be a challenge.*" These are challenges that some SSDs are helping solve [13]. Finally, the downsides of auditory feedback led GS to suggest "*Tactile information... fastened to your wrist [which] would vibrate... that's a lot easier to handle.*" These possibilities have also been recently explored [48].

Users want a fine degree of control over what information is translated into sound. In particular, for devices with wide field of views, users desired the ability to systematically reduce this or select key areas, in order to focus on specific information [9]. This sense of adjustability to personal needs provides not only control over the tool, but over their environment and experiences. Empowering users to solve their own accessibility problems [24] needs to be fostered in the design of SSDs to both solve practical problems as well as enhance everyday dialogues and social integration.

Users with impaired vision or visual prostheses such as the Argus II, can also benefit from using SSDs, e.g. visual gaze has been shown to be influenced by SSDs [47], furthermore the visual resolution of SSDs can exceed the Argus II [17, 49]. As such, SSDs can play a complimentary role in learning to use visual prostheses, or offsetting weaknesses.

Beyond practical needs and towards visual curiosity

One of the biggest drawbacks of many of the devices was the amount of time and effort it takes to learn them. Whilst the development of better training would help [26], the users were more interested in developing intuitive-to-use devices, which are not only efficient to use when needed but also provide aesthetically pleasing experiences. The interviewees also expressed an interest in learning about features of the world typically experienced by vision (e.g. the colour of the sky, the stars, and the Eiffel tower). This extends to sensing new aspects of their environment or reminiscing about sensations no longer accessible to them. SSDs do not yet cover this design space, however considering the interest in digitising visual information about the world and beyond (inspired by increased interest in big data and internet of things), there is a wealth of meaningful experiences translatable by SSDs for blind people to experience and participate in.

Limitations

While the present study provides a key perspective on the advancement of SSDs, there were several limiting factors that should be considered. First is that only vision into sound was investigated, as such many of our findings may not be applicable to visual to tactile translations, different substitutions (e.g. audio into tactile [29, 42]) or combinations (e.g. vision into sound and touch). While qualitative studies provide an overview of the concerns and desires of potential users, future quantitative investigations are required to refine these ideas and meet the expectations of users. Furthermore, different groups of users may have different desires, attributable to their visual experiences (e.g. congenitally blind, low vision) or expertise. Since we interviewed only ten BVIPs, future studies could further explore these sub-groups. Nevertheless our findings provide a strong motivation to go beyond current SSD designs and adopt a more 'human-centred' approach that puts their experiences at the forefront to enable and empower users.

CONCLUSION

This paper introduced three novel visual-to-auditory SSDs (Depth-vOICe, Synaestheatre, Creole) that represent a large range of potential design choices. Those devices were used as a technology probe by potential BVI end users in order to identify how these devices could address their personal challenges and visual aspirations. Interviews identified key problems and desires for such devices, placing value on specific visual features and auditory aesthetics. Designing SSDs with the end user's goals in mind should enable access to valuable experiences, solve problems and enrich everyday life. This experience-centered approach is unique within the wider sensory substitution field and provides key insights for interaction design. As one of our interviewees put it "*...it's not sight specific, it's human specific*."

ACKNOWLEDGEMENTS

This work was funded by the RM Phillips Foundation and supported by the EC within the Horizon2020 programme through the ERC (StG 638605). We thank Patrick Holroyd, Matt Martin, and Daniel Hajas for their great help in producing the accompanying video. Special thanks go to our participants for their time and valuable insights.

REFERENCES

1. @SEEINGWITHSOUND. 2015. As of last Thursday, The vOICe for Android has been installed by 269,395 unique users via Google Play https://play.google.com/store/apps/details?id=vOICe.vOICe … #blind #a11y. Tweet. (13 June, 2015). Retrieved September 21, 2015 from https://twitter.com/seeingwithsound/status/609623866319720448
2. @SEEINGWITHSOUND. 2015. Also during 2015: 21% of all runs of The vOICe for Android is by blind users! Get it on Google Play at https://t.co/XZKnjZMjZc #blind #a11y. Tweet. (13 June, 2015). Retrieved September 21, 2015 from https://twitter.com/seeingwithsound/status/609634973113556992
3. Sami Abboud, Shlomi Hanassy, Shelly Levy-Tzedek, Shachar Maidenbaum and Amir Amedi. 2014. EyeMusic: Introducing a "visual" colorful experience for the blind using auditory sensory substitution. *Restor Neurol Neurosci* 32, 2: 247-257.
4. Amir Amedi, William M Stern, Joan A Camprodon, Felix Bermpohl, Lotfi Merabet, Stephen Rotman, Christopher Hemond, Peter Meijer and Alvaro Pascual-Leone. 2007. Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nat Neurosci* 10, 6: 687-689.
5. Malika Auvray, Sylvain Hanneton, and J Kevin O'Regan. 2007. Learning to perceive with a visuo-auditory substitution system: Localisation and object recognition with 'The vOICe.' *Perception* 36, 3: 416-430.
6. Guido Bologna, Benoît Deville, Thierry Pun and Michel Vinckenbosch. 2007. Transforming 3D coloured pixels into musical instrument notes for vision substitution applications. EURASIP *J Image Video Process* 2: 8-8.
7. David J. Brown, Andrew J. R. Simpson and Michael J. Proulx. 2014. Visual objects in the auditory system in sensory substitution: how much information do we need? *Multisens Res* 27, 5-6: 337-57.
8. David J. Brown, Tom Macpherson, and Jamie Ward. 2011. Seeing with sound? Exploring different characteristics of a visual-to-auditory sensory substitution device. *Perception* 40, 9: 1120-1135.
9. Galit Buchs, Shachar Maidenbaum, Shelly Levy-Tzedek and Amir Amedi. 2015. Virtually Zooming-In with Sensory Substitution for Blind Users. Retrieved September 21, 2015 from http://www.rehabweek.org/downloads/ICVR_Buchs.pdf
10. Sandrine Ceurstemont. 2015. Substituting senses lets blind people take sonic holiday snaps. Retrieved September 21, 2015 from https://www.newscientist.com/article/dn28040-substituting-senses-lets-blind-people-take-sonic-holiday-snaps/
11. Daniel-Robert Chebat, Fabien C. Schneider, Ron Kupers and Maurice Ptito. 2011. Navigation with a sensory substitution device in congenitally blind individuals. *Neuroreport* 22, 7: 342-347.
12. Karla K. Evans and Anne Treisman. 2010. Natural cross-modal mappings between visual and auditory features. *J Vision* 10, 1: 6.
13. Alexander Fiannaca, Ilias Apostolopoulous, and Eelke Folmer. 2014. Headlock: A wearable navigation aid that helps blind cane users traverse large open spaces. In *Proc 16th int ACM SIGACCESS conf on Comp Acc*, 19-26.
14. Pat Fletcher. 2002. Seeing with sound: A journey into sight. Retrieved September 21, 2015 from http://www.seeingwithsound.com/tucson2002.html
15. Reginald Golledge, Roberta Klatzky, Jack Loomis and James Marston. 2004. Stated preferences for components of a personal guidance system for nonvisual navigation. *J Visual Imp Blind* 98, 3: 135-147.
16. José Luis González-Mora, Antonio F. Rodriguez-Hernandez, Enrique Burunat, F. Martin and Miguel A. Castellano. 2006. Seeing the world by hearing: Virtual Acoustic Space (VAS) a new space perception system for blind people. In *Info Comm Tech*, *ICTTA'06*, 2, 1: 837-842.
17. Alastair Haigh, David J. Brown, Peter Meijer, and Michael J. Proulx. 2013. How well do you see what you hear? The acuity of visual-to-auditory sensory substitution. *Front Psychol,* 4: 1-13.
18. Giles Hamilton-Fletcher and Jamie Ward. 2013. Representing colour through hearing and touch in sensory substitution devices. *Multisens Res* 26, 6; 503-32.
19. Giles Hamilton-Fletcher and Jamie Ward. 2014. What does red sound like? Establishing and applying aesthetics to sensory substitution. 2014. Retrieved September 21, 2015 from https://www.researchgate.net/publication/280640062_What_does_red_sound_like_Establishing_and_applying_aesthetics_to_sensory_substitution
20. Giles Hamilton-Fletcher, Thomas D. Wright, and Jamie Ward. 2015. Cross-modal correspondences enhance performance on a colour-to-sound sensory substitution device. Retrieved January 3rd, 2016 from

https://www.researchgate.net/publication/285673654_Cross-Modal_Correspondences_Enhance_Performance_on_a_Colour-to-Sound_Sensory_Substitution_Device

21. Intel® RealSense™ 3D Camera. 2015. Retrieved September 21, 2015 from http://www.intel.com/content/www/us/en/architecture-and-technology/realsense-depth-camera.html
22. Ali Israr, Olivier Bau, Seung-Chan Kim and Ivan Poupyrev. 2012. Tactile feedback on flat surfaces for the visually impaired. In *CHI 12 Ext Abs Hum Fac Comp Sys* 1571-1576.
23. Jung-Kyong Kim and Robert J. Zatorre. 2008. Generalized learning of visual-to-auditory substitution in sighted individuals. *Brain Res* 1242: 263-275.
24. Richard E. Ladner. 2015. Design for user empowerment. *Interactions* 22, 2: 24-29.
25. Jörg, Lewald. 2002. Vertical sound localization in blind humans. *Neuropsychologia*, 40, 12:1868-1872.
26. Shachar Maidenbaum, Sami Abboud and Amir Amedi. 2014. Sensory substitution: closing the gap between basic research and widespread practical visual rehabilitation. *Neurosci Biobehav Rev* 41: 3-15.
27. Olena Markaryan. 2015. Seeing with sound: what hidden abilities do we have? Retrieved September 21, 2015 from http://bionic-eyes-arms-eng.blogspot.co.uk/2015/09/seeing-with-sound-what-hidden-abilities.html
28. Lawrence E. Marks. 1987. On cross-modal similarity: Auditory–visual interactions in speeded discrimination. *J Exp Psy: Hum Percept Perf* 13, 3: 384-394.
29. Luke R. Mason. 2015. Rewiring the brain to create new senses. Retrieved September 21, 2015 from http://motherboard.vice.com/read/rewiring-the-brain-to-create-new-senses
30. Peter B. Meijer. 1992. An experimental system for auditory image representations. *IEEE Trans Biomed Eng,* 39, 2: 112-21.
31. Lofti, B. Merabet, Lorella Battelli, Souzana Obretenova, Sara Maguire, Peter Meijer, and Alvaro Pascual-Leone. 2009. Functional recruitment of visual cortex for sound encoded object identification in the blind. *Neuroreport* 20, 2: 132.
32. Project Tango – Google. 2015. Retrieved September 21, 2015 from https://www.google.com/atap/project-tango/
33. Michael J. Proulx, Petra Stoerig, Eva Ludowig, and Inna Knoll. 2008. Seeing 'Where' through the Ears: Effects of Learning-by-Doing and Long-Term Sensory Deprivation on Localization Based on Image-to-Sound Substitution. *Plos One,* 3, 3: e1840.
34. Brigitte Roèder, Wolfgang Teder-Saèlejaèrvi, Anette Sterr, Frank Roèsler, Steven A. Hillyard and Helen J. Neville. 1999. Improved auditory spatial tuning in blind humans. *Nature* 400, 6740: 162-166.
35. Hervé Segond, Déborah Weiss and Eliana Sampaio. 2005. Human spatial navigation via a visuo-tactile sensory substitution system. *Perception* 34, 10: 1231-1249.
36. Charles Spence. 2011. Crossmodal correspondences: A tutorial review. *Atten Percept Psychophy* 73, 4: 971-995.
37. Ella Striem-Amit, Laurent Cohen, Stanislas Dehaene, and Amir Amedi. 2012. Reading with sounds: sensory substitution selectively activates the visual word form area in the blind. *Neuron* 76, 3: 640-652.
38. Ella Striem-Amit, Miriam Guendelman and Amir Amedi. 2012. 'Visual' Acuity of the Congenitally Blind Using Visual-to-Auditory Sensory Substitution. *Plos One,* 7, 3: 1-6.
39. M. Iftekhar Tanveer, A. S. M. Iftekhar Anam, Mohammed Yeasin and Majid Khan. 2013. Do you see what I see?: designing a sensory substitution device to access non-verbal modes of communication. In *Proc 15th Int ACM SIGACCESS Conf Com Acc* 10.
40. TapTapSee - Blind and Visually Impaired Camera. Retrieved September 21, 2015 from http://www.taptapseeapp.com/
41. Bijal Trivedi. 2010. Sensory hijack: rewiring brains to see with sound. Retrieved September 21, 2015 from https://www.newscientist.com/article/mg20727731-500-sensory-hijack-rewiring-brains-to-see-with-sound/
42. Steven Wall and Stephen Brewster. 2006. Feeling what you hear: Tactile feedback for navigation of audio graphs. In *Proc SIGCHI conf Hum Fac Com Sys* 1123-1132.
43. Jamie Ward and Peter Meijer. 2010. Visual experiences in the blind induced by an auditory sensory substitution device. *Consci Cog* 19, 1: 492-500.
44. Jamie Ward and Thomas Wright. 2014. Sensory Substitution Devices as Advanced Sensory Tools. Retrieved September 21, 2015 from http://www.researchgate.net/profile/Jamie_Ward/publication/266315050_Sensory_Substitution_Devices_as_Advanced_Sensory_Tools/
45. Michele A. Williams, Caroline Galbraith, Shaun K. Kane, and Amy Hurst. 2014. "Just let the cane hit it": how the blind and sighted see navigation differently. In *Proc 16th int ACM SIGACCESS conf on Comp Acc*, 217-224.
46. Michael Wilson. 2015. “Who put that there!” The barriers to blind and partially sighted people getting out and about. Retrieved September 21, 2015 from https://www.rnib.org.uk/sites/default/files/Who%20put%20that%20there!%20Report%20February%202015.pdf

47. Thomas D. Wright, Jamie Ward, Sarah Simonon, and Aaron Margolis. 2012. "Where's Wally? Audio–visual mismatch directs ocular saccades in sensory substitution. *See Perc* 25: 61-61.

48. Hanlu Ye, Meethu Malu, Uran Oh, and Leah Findlater. 2014. Current and future mobile and wearable device use by people with visual impairments. In *Proc SIGCHI conf Hum Fac Com Sys* 3123-3132.

49. David D. Zhou, Jessy D. Dorn, and Robert J. Greenberg. 2013. The Argus® II retinal prosthesis system: An overview. In *IEEE Inter Conf Multimed Expo Works (ICMEW)*, pp. 1-6.